



ENVIRONMENTAL
HEALTH
PERSPECTIVES

<http://www.ehponline.org>

Identification of DNA Methylation Changes in Newborns Related to Maternal Smoking during Pregnancy

Christina A. Markunas, Zongli Xu, Sophia Harlid,
Paul A. Wade, Rolv T. Lie, Jack A. Taylor, and Allen J. Wilcox

<http://dx.doi.org/10.1289/ehp.1307892>

Received: 15 November 2013

Accepted: 4 June 2014

Advance Publication: 6 June 2014

Identification of DNA Methylation Changes in Newborns Related to Maternal Smoking during Pregnancy

Christina A. Markunas,^{1*} Zongli Xu,^{1*} Sophia Harlid,² Paul A. Wade,² Rolv T. Lie,^{3,4} Jack A. Taylor,^{1,2} and Allen J. Wilcox¹

¹Epidemiology Branch, National Institute of Environmental Health Sciences, NIH, DHHS, Research Triangle Park, North Carolina, USA; ²Laboratory of Molecular Carcinogenesis, National Institute of Environmental Health Sciences, NIH, DHHS, Research Triangle Park, North Carolina, USA; ³Department of Global Public Health and Primary Care, University of Bergen, Norway; ⁴Medical Birth Registry of Norway, Norwegian Institute of Public Health, Bergen, Norway. *These authors contributed equally to this work.

Address correspondence to Allen J. Wilcox, National Institute of Environmental Health Sciences, P.O. Box 12233, Mail Drop A3-05, Research Triangle Park, North Carolina 27709 USA. Telephone: 1 (919) 541-4660. Fax: 1 (919) 541-2511. E-mail: wilcox1@niehs.nih.gov

Running title: Smoking related methylation changes in newborns

Acknowledgments: We would like to thank all individuals for participating in the Norway Facial Clefts Study. In addition, we would like to thank Drs. Joubert and Wilson, as well as three anonymous reviewers for critically reviewing our manuscript. This research was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences.

Competing financial interests: The authors declare they have no actual or potential competing financial interests.

Abstract

Background: Maternal smoking during pregnancy is associated with significant infant morbidity and mortality, and may influence later disease risk. One mechanism by which smoking (and other environmental factors) might have long-lasting effects is through epigenetic modifications such as DNA methylation.

Objectives: To conduct an epigenome-wide association study (EWAS) investigating alterations in DNA methylation in infants exposed *in utero* to maternal tobacco smoke, using the Norway Facial Clefts Study.

Methods: The Illumina HumanMethylation450 BeadChip was used to assess DNA methylation in whole blood from 889 infants shortly after delivery. Out of 889 mothers, 287 reported smoking – twice as many smokers as in any previous EWAS of maternal smoking. CpG sites related to maternal smoking during the first trimester were identified using robust linear regression.

Results: We identified 185 CpGs with altered methylation in infants of smokers at genome-wide significance ($q\text{-value} < 0.05$; Mean $\Delta\beta = \pm 2\%$). These correspond to 110 gene regions, of which 7 have been previously reported and 10 are newly confirmed using publically-available results. Among these 10 the most noteworthy are *FRMD4A*, *ATP9A*, *GALNT2*, and *MEG3*, implicated in processes related to nicotine dependence, smoking cessation, and placental and embryonic development.

Conclusions: Our study identified 10 genes with newly established links to maternal smoking. Further, we note differences between smoking-related methylation changes in newborns and adults, suggesting possible distinct effects of direct versus indirect tobacco-smoke exposure as well as potential differences due to age. Further work would be needed to determine if these

small changes in DNA methylation are biologically or clinically relevant. The methylation changes identified in newborns may mediate the association between *in utero* maternal smoking exposure and later health outcomes.

Introduction

Cigarette smoke contains over 7000 chemicals, of which hundreds are known to be harmful and at least 69 are known to cause cancer (CDC 2010). Although the significant health effects of smoking are well recognized, smoking remains the largest preventable cause of death in the United States (CDC 2010). In addition to the deleterious direct effects on the person who smokes, smoking can also have indirect effects, for example, on the developing embryo or fetus. Maternal smoking during pregnancy is associated with substantial infant morbidity and mortality (Dietz et al. 2010). While smoking is a modifiable risk behavior that influences the health of both the infant and the mother, only about 45% of women who report that they smoke three months prior to pregnancy quit during pregnancy (CDC 2012). Exposure to the fetus thus remains an important area of research, with little known about the molecular changes that occur in newborns in response to exposure *in utero*. Further, if these changes persist for an extended period of time (or are permanent), they could have implications for disease development later in life.

The “developmental origins of disease” hypothesis (Gluckman and Hanson 2004) proposes that environmental exposures occurring during development can cause biological changes that influence disease susceptibility later in life. One potential mechanism is through an alteration of the fetal epigenome, which may have long-term consequences (Waterland and Michels 2007). One of the most well studied forms of epigenetic modification is DNA methylation, which occurs predominately at position C5 of cytosine in cytosine-guanine dinucleotides (CpGs). DNA methylation plays an important role in human health and has been associated with a growing number of diseases including cancer, imprinting disorders, and repeat-instability diseases (Robertson 2005).

Several studies have examined DNA methylation changes associated with indirect maternal tobacco smoke exposure *in utero* (reviewed in (Suter et al. 2013)). These studies have examined changes in DNA from placenta, cord blood, buccal cells, or granulocytes by taking either a targeted (i.e. one or a few genes) or a more global approach (i.e. epigenome-wide). However, only one study to date has used a high-density DNA methylation array (the Illumina HumanMethylation450 BeadChip), to assess DNA methylation in newborns related to maternal smoking during pregnancy (Joubert et al. 2012).

In order to further explore DNA methylation changes in infants related to maternal smoking during pregnancy, we performed an epigenome-wide association study (EWAS) using whole blood from 889 infants in the Norway Facial Clefts Study (NCL), including 287 whose mothers smoked during the first trimester. As a replication set, we used publically available results (Obtained from Supplemental Table S1 in (Joubert et al. 2012)) and considered two levels of replication: 1) Site-level at the exact CpG and 2) Gene-level for other CpG sites in or near the same gene. In addition, we considered the overlap between smoking-associated methylation in adults and newborns, with special focus on genes that may be unique to each.

Methods

Study population

The present study is based on infants from the Norway Facial Clefts Study (NCL). NCL is a national population-based case-control study of cleft lip and cleft palate, disorders characterized by the incomplete fusion of the lip and/or palate during development. The study design has been previously described in detail (Wilcox et al. 2007). Briefly, between the years of 1996 and 2001 all families of newborns referred for cleft surgery in Norway were contacted; 88% of those eligible agreed to participate (N=573). Controls were selected by a random sampling of roughly

4 per 1000 live births in Norway during that same time period; 76% of those eligible agreed to participate (N=763). NCL was approved by the Norwegian Data Inspectorate and Regional Medical Ethics Committee of Western Norway and informed consent was provided by both the mother and father. In the current study, a subset of 418 facial cleft cases (Lip only=107, Palate only=144, Lip and palate=167) and 480 controls were selected based on DNA availability. Newborn blood was collected at the delivery hospital in plain glass capillary tubes by heel stick, then frozen and shipped to a central laboratory in Oslo for phenylketonuria and other routine genetic analysis. Collection was identical for cases and controls. Study population characteristics are shown in Tables 1 and 2.

Tobacco smoke exposure

Information about maternal tobacco smoke exposure during pregnancy was obtained through mother's questionnaire around 3-4 months after delivery. An English translation of this questionnaire can be found on the study's website (NIEHS 2008). Mothers were asked about active smoking during the first trimester (average number of cigarettes smoked per day or per month) and passive smoke exposure at any time during the pregnancy (average number of hours per day she was within 2 meters of a smoker). Environmental tobacco smoke exposure (passive smoking) *in utero* was not associated with DNA methylation in newborns (False discovery rate (FDR) $q < 0.05$), and so we combined non-smokers, passive (≥ 1 hour/day within 2 meters of a smoker), and infrequent smokers (< 1 cigarette per day) into one group (N=607). All active smokers were also combined (N=290). We explored the possibility of a dose response relationship between the extent of smoke exposure and the level of methylation at genome-wide significant CpG sites (FDR $q < 0.05$). Three exposure classes were defined for this analysis: 1) 607 non-active smokers: non-smokers (N=437), passive (N=131), and infrequent smokers

(N=39), 2) 141 light smokers (1-5 cigarettes per day), and 3) 149 moderate (6-10 cigarettes per day; N=107) and heavy smokers (≥ 11 cigarettes per day; N=42). In order for a CpG site to meet our criteria for a dose-response relationship it must satisfy two requirements: 1) a mean methylation level that progressively increases or decreases across the three categories of smoking and 2) the 3-level smoking variable must be significant (FDR $q < 0.05$) when treated as a continuous variable in the robust linear regression model. One mother used snuff tobacco; her baby was excluded from all analyses.

Analysis

A description of DNA methylation data generation (Illumina HumanMethylation450), quality assessment, and pre-processing (e.g. normalization) are provided (See Supplemental Material, Methods: Data generation, Data cleaning and quality assessment, and Data pre-processing). After exclusions (See Supplemental Material, Methods: Data cleaning and quality assessment), 889 samples ($N_{\text{smokers}} = 287$, $N_{\text{non-smokers}} = 602$) and 357,320 CpG probes remained for analysis. A robust linear regression model was employed using the R package, MASS (Venables and Ripley 2002), to test the association between maternal smoking status (binary variable: active smokers versus non-smokers and passive/infrequent smokers) and methylation (β -value) at each CpG site, adjusting for facial cleft status (control, cleft lip with or without cleft palate and cleft palate only), infant's sex, and two technical factors: batch effects (96-well plate) and bisulfite conversion efficiency (Model1) (See Supplemental Material, Methods: Data pre-processing). A sensitivity analysis was also conducted adjusting for several potential confounders. Maternal alcohol consumption, education, age, body mass index (BMI), dietary folate, folic acid supplement use, multivitamin use, parity, gestational age, and infant's birth weight were all considered as possible confounders. Only those associated with maternal smoking ($p < 0.05$)

(Tables 1 and 2) and that could potentially influence DNA methylation were added to the secondary model, referred to below as “Model2”. Cellular heterogeneity is a potential confounder if the distribution of blood cell subtypes differs by maternal smoking status. In order to correct for cellular heterogeneity, we implemented the method developed by Houseman and colleagues (Houseman et al. 2012) to estimate five different cell proportions for each sample and then adjust for those in our robust linear regression model (Model2). A reference dataset ((Reinius et al. 2012), GSE35069) was used first to identify DNA methylation profiles that were specific to each of the five cell subtypes. This information was then used to generate sample-specific cell proportion estimates that were included in Model2 to correct for five leukocyte subtype proportions (See Supplemental Material, Table S1 for a comparison between newborns exposed and unexposed). To correct for multiple testing, we estimated the FDR using the q-value framework (Storey and Tibshirani 2003).

Results

Epigenome-wide association study

Epigenome-wide association results based on Model1 are presented in Figure 1. In addition, a volcano plot and quantile-quantile (Q-Q) plot are provided as supplemental material (Figures S1-S2). Further adjustment for potential confounders in Model2 (maternal alcohol use, maternal education, maternal age at delivery, maternal folic acid supplement use, infant’s birth weight, and 5 different leukocyte cell type proportions) did not substantively affect the results as assessed by examining the correlation between the p-values and beta coefficients across all CpGs tested (see Supplemental Material, Figure S3). The more parsimonious model, Model1, was therefore selected as the primary model. Out of 357,320 CpG sites tested for association with maternal smoking, we identified 185 CpG sites with altered methylation at genome-wide

significance (FDR $q < 0.05$) (Tables S2-S3) found within 110 gene regions. The term, ‘gene region’, was used to annotate CpGs located within or outside the transcribed region of a gene. The ‘gene region’ for each CpG probe was defined as the nearest gene if the CpG site was intergenic or the gene itself if the CpG site was located within the gene body (see Supplemental Material, Table S3). Comparing exposed and unexposed infants, 43% of the 185 CpGs had decreased methylation (negative beta coefficient) and 57% had increased methylation (positive beta coefficient). The average percentage change in methylation ($\text{Mean } \beta_{\text{smokers}} - \text{Mean } \beta_{\text{non-smokers}}$) was small, ranging from -8% to +7% with a mean of -2% for sites with decreased methylation and +2% for sites showing increased methylation. Further annotation of the significant CpG sites, including distance to the closest transcription start site (TSS) and whether it is located within a CpG island, shelf, or shore is provided (see Supplemental Material, Table S4). We compared the distribution of the 185 differentially methylated CpG sites among genomic regions (CpG island, shelf, shore, or non-island), using the Illumina Human Methylation 450 BeadChip as a reference (see Supplemental Material, Table S5). In particular, we observed a significant enrichment for CpG shores when we examined all 185 CpGs (Chi-square test, $p=0.002$); the same thing was observed when the analysis was restricted to those 80 sites with decreased methylation (Chi-square test, $p=0.003$).

The relationship between the extent of smoke exposure and the level of methylation was also assessed. Out of the 185 genome-wide significant CpG sites, only 107 CpGs (58%) showed evidence of a dose-response relationship (see Supplemental Material, Figure S4 and Table S6). When our results were restricted to the 23 novel CpG sites confirmed at the site or gene-level (Table 3), 18 of the sites (78%) showed a dose-response relationship. Expanding that list of 23 CpGs to include all 62 CpG sites confirmed at the site or gene-level, 49 (79%) met the criteria

for a dose-response relationship. Further, CpGs with the smallest p-values had the highest proportion of dose-response relationships (see Supplemental Material, Table S7), with 91% of the CpGs showing a dose-response relationship with $p < 1 \times 10^{-10}$.

Replication using publically available EWAS results

We identified 185 CpGs (corresponding to 110 gene regions) with altered methylation at genome-wide significance (FDR $q < 0.05$; see Supplemental Material, Tables S2-S3). In order to replicate our findings, we used results from the only previous maternal smoking study using Illumina HumanMethylation450 BeadChips (Results obtained from Supplemental Table S1 in (Joubert et al. 2012); see Supplemental Material, Table S8). In that study, which used a conservative Bonferroni correction for 473,844 tests ($p < 1.06 \times 10^{-7}$) as the criterion for genome-wide significance, 26 CpGs were reported as genome-wide significant. We first describe our confirmation of those 26 CpGs (10 gene regions) that we further classify as either established (N=5; identified previously at genome-wide significance and replicated using an independent population) or implicated (N=21; identified previously at genome-wide significance, currently lacking replication). We confirmed the 5 out of 26 CpG sites (3 gene regions: *AHRR*, *CYP1A1*, and *GFII*) that were additionally replicated (Bonferroni correction for 26 tests: $p < 0.0019$) by Joubert and colleagues using an independent population of infants born to 18 smoking and 18 non-smoking mothers from the Newborn Epigenetics Study (NEST) (Figure 2A, (Joubert et al. 2012)). In addition, we replicated 17 of the 21 CpGs that were previously identified at genome-wide significance, but not directly replicated using 36 mothers from NEST ($p < 0.0019$) although had replication p-values ranging from 0.002 to 0.066 (Figure 2A). Replication of those 17 sites established an additional 3 gene regions (*MYO1G*, *CNTNAP2*, *LOC100507468*) related to maternal smoking during pregnancy. In addition to the 22 CpGs replicated as described above,

an additional 17 CpGs, located in 7 of the 10 previously implicated or established gene regions met genome-wide significance in our study, bringing the total number to 39 CpGs in 7 gene regions that were confirmed at either a site (22 CpGs with exact CpG site confirmation; Figure 2A) or gene-level (17 CpGs without exact CpG site confirmation but in or near one of the 10 genes). In addition to the 26 maternal smoking-related CpGs previously reported, Joubert and colleagues provided results for 74 nominally significant CpGs (part of their top 100 findings) representing 69 gene regions that did not meet their strict criteria for genome-wide significance (Bonferroni correction, $p < 1.06 \times 10^{-7}$). We used this set of results to serve as an independent replication set for our novel findings (146 CpGs corresponding to 103 gene regions not previously implicated (Figure 2B; see Supplemental Material, Table S8)), which we next describe in detail below.

Using the publically available results consisting of 74 nominally significant CpGs we find confirmation of 9 CpGs in 8 gene regions (*GNG12*, *LOC284998*, *VGLL4*, *FRMD4A*, *CUX2*, *MEG3*, *FTO*, and *ATP9A*). Further, we identified genome-wide significant CpGs in 2 additional gene regions (*GALNT2* and *KIF26B*) where a different CpG site in the same gene region showed altered DNA methylation (i.e. the exact CpG sites were not identical). In total, we identified 8 CpGs in these 2 gene regions and 15 CpGs in the 8 gene regions that met genome-wide significance in our study (FDR $q < 0.05$, Figure 2B, Table 3). This set of 10 genes is the focus of the discussion below. Five of these 10 gene regions, *GNG12*, *GALNT2*, *FRMD4A*, *MEG3*, and *FTO* contain multiple differentially methylated CpG sites at genome-wide significance. Notably, all confirmed sites showed the same direction of effect in both studies. The remaining 123 CpGs (representing 93 gene regions) that achieved genome-wide significance (FDR $q < 0.05$) in our study require further confirmation.

Smoking-related methylation changes in adults and newborns

We also considered our results in relation to three published studies of adult smoking that used the Illumina HumanMethylation450 BeadChip for their primary analysis (Monick et al. 2012; Shenker et al. 2012; Zeilinger et al. 2013). The CpG sites related to adult smoking include: 1) 187 genome-wide significant sites ((Zeilinger et al. 2013); $p \leq 1 \times 10^{-7}$; obtained from their Table S2), 2) Top findings for breast ($N_{\text{sites}}=17$) and colon cancer ($N_{\text{sites}}=19$) EWASs of adult smoking ((Shenker et al. 2012); $p < 1 \times 10^{-5}$; obtained from their Table S3), and 3) Top 30 differentially methylated CpGs identified in lymphoblasts and alveolar macrophages ((Monick et al. 2012); some, but not all are genome-wide significant; obtained from their Tables 2 and 5). Using these lists, we examined the overlap between smoking-related DNA methylation in adults and newborns. Out of the 185 genome-wide significant CpG sites identified in our study, 21 CpGs (6 gene regions) were identified with the same direction of effect in at least one of the three adult smoking studies (see Supplemental Material, Table S8).

We next sought to more generally compare the established smoking-related genes in adults and newborns (see Supplemental Material, Table S8). Ten gene regions were identified as among the top findings in at least two out of the three adult studies. We compared the two newborn studies using gene regions from the 185 CpGs in our study and the 100 CpGs reported by Joubert, et al. and found 17 gene regions in common. Note that in both the adult and newborn comparisons, the same gene region was implicated but the exact CpG site may differ between studies. Of the 10 established gene regions identified in multiple adult studies, 4 were identified in both newborn studies (*AHRR*, *GNG12*, *GFII*, *CNTNAP2*). The remaining six gene regions (*LOC100131546*, *TMEM51*, *IER3*, *GNAI2*, *LRP5*, *F2RL3*) were restricted to adults only. In contrast, of 17 established gene regions identified in both newborn studies, 6 were reported in at least one adult

study (*AHRR*, *CNTNAP2*, *GNG12*, *GFII*, *ATP9A*, *MYO1G*), while 11 (*LOC284998*, *FTO*, *CUX2*, *CYP11A1*, *VGLL4*, *KIF26B*, *MEG3*, *FRMD4A*, *GALNT2*, *LOC100507468*, *EXT1*) were restricted to newborns only.

Our final comparison between adult and newborn findings was related to the direction of methylation change. In contrast to newborns, where 43% of CpGs had decreased methylation, over 70% of the CpGs identified in adult smokers had decreased methylation. However, there is some limited evidence, at least in adults, that this finding may be tissue specific (Monick et al. 2012).

Discussion

We conducted an EWAS investigating alterations in DNA methylation in infants indirectly exposed *in utero* to maternal tobacco smoke. We identified 185 CpG sites that met genome-wide significance (FDR $q < 0.05$). Publically-available results from a recent EWAS of maternal smoking were used to confirm some of our findings (Joubert et al. 2012). We also describe similarities and differences between smoking-related methylation changes in adults and newborns.

Our study confirmed 5 of 5 previously replicated maternal smoking-related CpG sites representing 3 gene regions. We also replicated another 17 of 21 CpGs, representing 3 additional gene regions, that had achieved genome-wide significance in the previous study but were not replicated in their small replication population using a Bonferroni correction for multiple tests (Joubert et al. 2012). Most of these CpGs, however, had p-values less than 0.05. Most importantly, we provide new findings of 8 gene regions associated with maternal smoking where, using the publically-available results from Joubert and colleagues, there was exact CpG

site confirmation and 2 gene regions where different CpGs showed evidence of association (Table 3). From this set of 10 genes with newly-established links to maternal smoking, the most noteworthy include *MEG3* (maternally expressed 3), *FRMD4A* (FERM domain containing 4A), *ATP9A* (ATPase, class II, type 9A), and *GALNT2* (UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 2), discussed in detail below.

In *MEG3* we find 2 genome-wide significant CpGs and 1 nominally associated CpG with increased methylation in infants of mothers who smoked (see Supplemental Material, Figure S5). *MEG3* is an imprinted gene (paternal allele silenced, maternal allele expressed) that encodes a long noncoding RNA implicated in tumorigenesis (Benetatos et al. 2011) and embryonic development (Zhou et al. 2010). Previous studies suggest a potential link between *MEG3*, tobacco smoke, and development. Expression of *MEG3* is down-regulated in immortalized human bronchial epithelial cells in response to treatment with cigarette smoke condensate but can be reactivated following 5-aza-2-deoxycytidine treatment (Hu et al. 2009). This suggests that the smoking-related decrease in expression may be associated with increased DNA methylation, consistent with the direction of effect found in newborns of mothers who smoke. Related to development, maternal cigarette smoking during pregnancy causes reduced birth weight (Suter et al. 2013) and expression of *MEG3* is down-regulated in placentae with intrauterine growth restriction (McMinn et al. 2006) as was observed in cigarette smoke condensate treated human bronchial epithelial cells. However, unlike the direction of effect observed in the previous study, *MEG3* is up-regulated in placentae of women smokers compared with nonsmokers (Votavova et al. 2011). Given these inconsistent findings, additional studies are needed to clarify this relationship.

The *DLK1/MEG3* imprinting domain on 14q32.2 has two differentially methylated regions (DMRs), the MEG3-DMR (secondary, post-fertilization derived DMR; located near our smoking-associated CpGs) and IG-DMR (primary, germline-derived intergenic DMR) (Skaar et al. 2012) that are hypothesized to act as imprinting control centers in the body and placenta, respectively (Kagami et al. 2010). Although the mechanism by which the MEG3-DMR regulates imprinting is not well understood, CTCF (CCCTC-binding factor) is thought to be involved (Skaar et al. 2012) perhaps by modulating chromatin boundary formation (Wylie et al. 2000). The *MEG3* smoking-associated CpGs identified in our study are located at the border of the MEG3-DMR immediately adjacent to two putative CTCF binding sites ((Wylie et al. 2000), see Supplemental Material, Figure S5). If increased methylation in newborns of smokers inhibits CTCF binding, it could result in decreased expression of *MEG3* and other maternally expressed genes, as well as increased expression of paternally expressed genes in that region (Kagami et al. 2010). Dysregulation of *MEG3* could potentially affect embryonic development and, if the altered methylation state persists for an extended period of time have long term health effects.

We identified altered methylation of CpGs near two genes, *FRMD4A* and *ATP9A*, associated with nicotine (Yoon et al. 2012) and substance dependence (Johnson et al. 2008), and the ability to quit smoking (Uhl et al. 2008). Increased methylation at 4 CpGs in newborns of smokers was observed in *FRMD4A*, a scaffolding protein implicated in the regulation of epithelial polarity (Ikenouchi and Umeda 2010). Decreased methylation with smoking exposure was found at 1 CpG in *ATP9A*, a class 2 P4-ATPase that localizes to endosomes and the trans-Golgi network (Takatsu et al. 2011). Although findings are not consistent across all studies, maternal smoking during pregnancy has been previously associated with smoking related traits in offspring (e.g. nicotine dependence) even after accounting for postnatal exposure to tobacco smoke (Al Mamun

et al. 2006; Kandel et al. 1994; Kardia et al. 2003; Lieb et al. 2003; Selya et al. 2012); however, the possibility cannot be ruled out that these associations are due to unmeasured confounding. An animal study provides further support for an association by showing a neurotoxicant effect of prenatal nicotine exposure in rats that persisted into adolescence (Abreu-Villaca et al. 2004). Authors suggested that if similar, persistent alterations occurred in offspring of mothers who smoked during pregnancy this could contribute to an increased susceptibility of nicotine dependence in adolescence. Thus, if the altered methylation states of *FRMD4A* and/or *ATP9A* persisted through adolescence, these could potentially contribute to the observed relationship between maternal smoking and smoking-related traits, such as nicotine dependence, in offspring.

In *GALNT2* we identified 7 CpGs at genome-wide significance that showed decreased methylation in newborns of smokers. *GALNT2* is associated with lipid levels (Teslovich et al. 2010), insulin signaling (Marucci et al. 2013a), hyperglycemia (Marucci et al. 2013b), and the malignant behavior of hepatocellular carcinoma (Wu et al. 2011). *GALNT2* may also play a role in extravillous trophoblast invasion, which is important for placental development (Liao et al. 2012). Aberrant DNA methylation of *GALNT2*, which has so far been identified only in newborns exposed and not adults, may affect the development of the placenta and could subsequently influence the health of the fetus.

In this study we describe some unique and common aspects of smoking-related methylation changes in adults and newborns. Less than half of the maternal-smoking-related genes that are now well-established in newborn studies were identified in any of the three published adult smoking studies. Moreover, some of the best-established adult-smoking-related gene regions (e.g. *F2RL3*) were not among the top results from either of the newborn studies. These observations suggest that there could be distinct effects of direct tobacco smoke exposure in

adults versus indirect tobacco smoke exposure *in utero*. There are several plausible explanations for this observation: 1) Age-related susceptibility: The epigenome of the developing embryo and fetus is thought to be particularly susceptible to environmental exposures as there are many epigenetic changes that are occurring along with rapid cell division (Perera and Herbstman 2011)., 2) Tissue specific alterations related to age at exposure: Exposure during development may be more likely to induce widespread changes (i.e. many tissues affected), whereas exposure during adulthood may be more likely to induce restricted changes, perhaps only identifiable in one or a few tissues (Rakyan et al. 2011)., 3) Exposure variability: The exposure itself could vary with respect to duration. In addition, the fetus is indirectly exposed to tobacco smoke, thus the actual exposure is dependent on a number of factors, including the ability of the placenta and the mother to metabolize and detoxify components of the cigarette smoke., and 4) Age-related differences in the response to exposure: Response to an environmental exposure, such as tobacco smoke, may vary by age due to differences in one's metabolism and capacity to detoxify the components of tobacco smoke (Aldridge et al. 2003). For example, in response to tobacco smoke, age-dependent expression (Eke et al. 1997) of genes involved in metabolizing components of tobacco smoke has been reported., and/or 5) Power of the studies: It is also possible that future newborn and adult studies with larger sample sizes will find more overlap of these genes.

There are several limitations of our study. We were able to confirm only some of our findings using publically available results. Our confirmations were based on the top 100 sites in a previously published study (Joubert et al. 2012), so we are unable to comment on the total number of CpGs that are currently lacking replication given the fact that we did not have access to a complete replication set. Even with this limited list we were able to identify 10 new genes

related to maternal smoking, including some with biologically plausible functions. Our correction for multiple testing was based on a reduced number of probes after filtering out the least variable 20% of CpGs. The alternative of filtering by degree of difference may induce FDR bias, but there is no evidence of such bias when filtering by variability (van Iterson et al. 2010). As a further check, we compared our results with and without filtering; the findings were essentially unchanged. Of the 23 novel CpG sites confirmed at the site or gene-level (Table 3), all but one remained significant at a genome-wide level without filtering (FDR $q < 0.05$), while the one exception (cg03687532) had an FDR $q = 0.06$. These additional results provide some reassurance about the robustness of our conclusions. Another potential limitation is that we used self-reported smoking as opposed to a more objective measure, such as cotinine. However, self-reported smoking is generally considered to be a valid measure of smoking. Self-reported maternal smoking and plasma cotinine concentrations were compared in the Mother and Child Cohort Study, another population-based Norwegian study conducted during 2002-2003. That study reported a high level of validity for the self-reported data (Kvalvik et al. 2012). In our study, mean birth weight was reduced with maternal smoking (Table 2), consistent with the usual expected effect of smoking on the fetus. If there was underreporting of smoking during pregnancy, this type of misclassification would only reduce our power, and would not explain our detection of 185 CpG sites at genome-wide significance. Furthermore, in the absence of further validation (e.g., pyrosequencing) and functional work, it is not known whether these changes are biologically relevant or which gene or genes they affect, if any. It also remains to be seen if these small changes in DNA methylation will be of clinical relevance.

Conclusions

Consistent with previous reports, we found maternal smoking during pregnancy to be associated with an altered DNA methylation profile in infants. Of particular interest were findings in *FRMD4A*, *ATP9A*, *GALNT2*, and *MEG3* – genes that are implicated in processes related to nicotine dependence, smoking cessation, and placental and embryonic development. Further, we describe some unique aspects of DNA methylation changes in adults compared with newborns. Direct versus indirect *in utero* smoke exposure appear to elicit differential epigenetic responses that may be due to age-related differences in susceptibility to tobacco smoke, tissue specific alterations related to age at exposure, variation in the exposure itself, and/or age-related differences in the response to exposure. It remains to be seen if the small changes in DNA methylation identified (Mean $\Delta\beta = \pm 2\%$) are biologically or clinically relevant. Future studies examining the persistence of DNA methylation changes due to an *in utero* exposure throughout life, and particularly associations of these changes with health phenotypes later in life (i.e. DNA methylation as a mediator of health effects) will be of particular interest.

References

- Abreu-Villaca Y, Seidler FJ, Slotkin TA. 2004. Does prenatal nicotine exposure sensitize the brain to nicotine-induced neurotoxicity in adolescence? *Neuropsychopharmacology* 29:1440-1450.
- Al Mamun A, O'Callaghan FV, Alati R, O'Callaghan M, Najman JM, Williams GM, et al. 2006. Does maternal smoking during pregnancy predict the smoking patterns of young adult offspring? A birth cohort study. *Tob Control* 15:452-457.
- Aldridge JE, Gibbons JA, Flaherty MM, Kreider ML, Romano JA, Levin ED. 2003. Heterogeneity of toxicant response: Sources of human variability. *Toxico Sci* 76:3-20.
- Benetatos L, Vartholomatos G, Hatzimichael E. 2011. Meg3 imprinted gene contribution in tumorigenesis. *Int J Cancer* 129:773-779.
- CDC. 2010. How tobacco smoke causes disease: The biology and behavioral basis for smoking-attributable disease, a report of the surgeon general, executive summary. Available: http://www.cdc.gov/tobacco/data_statistics/sgr/2010/ [accessed 16 July 2013].
- CDC. 2012. Pregnancy risk assessment monitoring system and smoking. Available: <http://www.cdc.gov/prams/tobaccoandprams.htm> [accessed 16 July 2013].
- Dietz PM, England LJ, Shapiro-Mendoza CK, Tong VT, Farr SL, Callaghan WM. 2010. Infant morbidity and mortality attributable to prenatal smoking in the U.S. *Am J Prev Med* 39:45-52.
- Eke BC, Vural N, Iscan M. 1997. Age dependent differential effects of cigarette smoke on hepatic and pulmonary xenobiotic metabolizing enzymes in rats. *Arch Toxicol* 71:696-702.
- Gluckman PD, Hanson MA. 2004. Developmental origins of disease paradigm: A mechanistic and evolutionary perspective. *Pediatr Res* 56:311-317.
- Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. 2012. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 13:86-102.
- Hu YC, Yang ZH, Zhong KJ, Niu LJ, Pan XJ, Wu DC, et al. 2009. Alteration of transcriptional profile in human bronchial epithelial cells induced by cigarette smoke condensate. *Toxicol Lett* 190:23-31.
- Ikenouchi J, Umeda M. 2010. Frmd4a regulates epithelial polarity by connecting Arf6 activation with the PAR complex. *Proc Natl Acad Sci U S A* 107:748-753.

- Johnson C, Drgon T, Liu QR, Zhang PW, Walther D, Li CY, et al. 2008. Genome wide association for substance dependence: Convergent results from epidemiologic and research volunteer samples. *BMC Med Genet* 9:113-123.
- Joubert BR, Haberg SE, Nilsen RM, Wang X, Vollset SE, Murphy SK, et al. 2012. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environ Health Perspect* 120:1425-1431.
- Kagami M, O'Sullivan MJ, Green AJ, Watabe Y, Arisaka O, Masawa N, et al. 2010. The IG-DMR and the MEG3-DMR at human chromosome 14q32.2: Hierarchical interaction and distinct functional properties as imprinting control centers. *PLoS Genet* 6:e1000992.
- Kandel DB, Wu P, Davies M. 1994. Maternal smoking during pregnancy and smoking by adolescent daughters. *Am J Public Health* 84:1407-1413.
- Kardia SL, Pomerleau CS, Rozek LS, Marks JL. 2003. Association of parental smoking history with nicotine dependence, smoking rate, and psychological cofactors in adult smokers. *Addict Behav* 28:1447-1452.
- Kvalvik LG, Nilsen RM, Skjaerven R, Vollset SE, Midttun O, Ueland PM, et al. 2012. Self-reported smoking status and plasma cotinine concentrations among pregnant women in the Norwegian mother and child cohort study. *Pediatr Res* 72:101-107.
- Liao WC, Chen CH, Liu CH, Huang MJ, Chen CW, Hung JS, et al. 2012. Expression of GALNT2 in human extravillous trophoblasts and its suppressive role in trophoblast invasion. *Placenta* 33:1005-1011.
- Lieb R, Schreier A, Pfister H, Wittchen HU. 2003. Maternal smoking and smoking in adolescents: A prospective community study of adolescents and their mothers. *Eur Addict Res* 9:120-130.
- Marucci A, Cozzolino F, Dimatteo C, Monti M, Pucci P, Trischitta V, et al. 2013a. Role of GALNT2 in the modulation of ENPP1 expression, and insulin signaling and action: GALNT2: A novel modulator of insulin signaling. *Biochim Biophys Acta* 1833:1388-1395.
- Marucci A, di Mauro L, Menzaghi C, Prudente S, Mangiacotti D, Fini G, et al. 2013b. GALNT2 expression is reduced in patients with type 2 diabetes: Possible role of hyperglycemia. *PLoS One* 8:e70159.

- McMinn J, Wei M, Schupf N, Cusmai J, Johnson EB, Smith AC, et al. 2006. Unbalanced placental expression of imprinted genes in human intrauterine growth restriction. *Placenta* 27:540-549.
- Monick MM, Beach SR, Plume J, Sears R, Gerrard M, Brody GH, et al. 2012. Coordinated changes in AHRR methylation in lymphoblasts and pulmonary macrophages from smokers. *Am J Med Genet B Neuropsychiatr Genet* 159B:141-151.
- NIEHS. 2008. Pregnancy, heredity, and environment questionnaire. Available: <http://www.niehs.nih.gov/research/atniehs/labs/epi/studies/ncl/question/> [accessed 7 July 2013].
- Perera F, Herbstman J. 2011. Prenatal environmental exposures, epigenetics, and disease. *Reprod Toxicol* 31:363-373.
- Rakyan VK, Down TA, Balding DJ, Beck S. 2011. Epigenome-wide association studies for common human diseases. *Nat Rev Genet* 12:529-541.
- Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D, et al. 2012. Differential DNA methylation in purified human blood cells: Implications for cell lineage and studies on disease susceptibility. *PLoS One* 7:e41361.
- Robertson KD. 2005. DNA methylation and human disease. *Nat Rev Genet* 6:597-610.
- Selya AS, Dierker LC, Rose JS, Hedeker D, Mermelstein RJ. 2012. Risk factors for adolescent smoking: Parental smoking and the mediating role of nicotine dependence. *Drug Alcohol Depend* 124:311-318.
- Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, et al. 2012. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. *Hum Mol Genet* 22:843-851.
- Skaar DA, Li Y, Bernal AJ, Hoyo C, Murphy SK, Jirtle RL. 2012. The human imprintome: Regulatory mechanisms, methods of ascertainment, and roles in disease susceptibility. *ILARJ* 53:341-358.
- Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 100:9440-9445.
- Suter MA, Anders AM, Aagaard KM. 2013. Maternal smoking as a model for environmental epigenetic changes affecting birthweight and fetal programming. *Mol Hum Reprod* 19:1-6.

- Takatsu H, Baba K, Shima T, Umino H, Kato U, Umeda M, et al. 2011. ATP9B, a P4-ATPase (a putative aminophospholipid translocase), localizes to the trans-Golgi network in a CDC50 protein-independent manner. *J Biol Chem* 286:38159-38167.
- Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. 2010. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466:707-713.
- Uhl GR, Liu QR, Drgon T, Johnson C, Walther D, Rose JE, et al. 2008. Molecular genetics of successful smoking cessation: Convergent genome-wide association study results. *Arch Gen Psychiatry* 65:683-693.
- van Iterson M, Boer JM, Menezes RX. 2010. Filtering, FDR and power. *BMC Bioinformatics* 11:450-461.
- Venables WN, Ripley BD. 2002. *Modern applied statistics with S*. 4th ed. New York:Springer.
- Votavova H, Dostalova Merkerova M, Fejglova K, Vasikova A, Krejcik Z, Pastorkova A, et al. 2011. Transcriptome alterations in maternal and fetal cells induced by tobacco smoke. *Placenta* 32:763-770.
- Waterland RA, Michels KB. 2007. Epigenetic epidemiology of the developmental origins hypothesis. *Annu Rev Nut* 27:363-388.
- Wilcox AJ, Lie RT, Solvoll K, Taylor J, McConaughy DR, Abyholm F, et al. 2007. Folic acid supplements and risk of facial clefts: National population based case-control study. *BMJ* 334:464-470.
- Wu YM, Liu CH, Hu RH, Huang MJ, Lee JJ, Chen CH, et al. 2011. Mucin glycosylating enzyme GALNT2 regulates the malignant character of hepatocellular carcinoma by modifying the EGF receptor. *Cancer Res* 71:7270-7279.
- Wylie AA, Murphy SK, Orton TC, Jirtle RL. 2000. Novel imprinted DLK1/GTL2 domain on human chromosome 14 contains motifs that mimic those implicated in IGF2/H19 regulation. *Genome Res* 10:1711-1718.
- Yoon D, Kim YJ, Cui WY, Van der Vaart A, Cho YS, Lee JY, et al. 2012. Large-scale genome-wide association study of asian population reveals genetic factors in FRMD4A and other loci influencing smoking initiation and nicotine dependence. *Hum Genet* 131:1009-1021.

Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. 2013. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PLoS One 8:e63812.

Zhou Y, Cheunsuchon P, Nakayama Y, Lawlor MW, Zhong Y, Rice KA, et al. 2010. Activation of paternally expressed genes and perinatal death caused by deletion of the Gtl2 gene. Development 137:2643-2652.

Table 1. Characteristics of mothers in the study according to maternal smoking.

Characteristic	Non-smokers	Smokers	P-value^a
All mothers	602	287	
Alcohol use ^b			4.7x10 ^{-4*}
0	410	158	
1-3	104	64	
4-6	37	20	
≥ 7	45	42	
Education			2.7x10 ^{-6*}
Less than high school	57	61	
High school and above	545	226	
Parity ^c			0.12
1	239	128	
2	209	102	
3	117	39	
≥ 4	37	18	
Age at delivery (years)	29.6 ± 4.9	28.0 ± 4.9	3.1x10 ^{-6*}
Pre-pregnancy BMI (kg/m ²)	23.6 ± 4.1	23.4 ± 4.3	0.56
Folic acid supplement (µg) ^d			0.02*
0	354	194	
1-399	134	58	
≥ 400	114	35	
Dietary folate (µg)			0.18
0-171	163	88	
172-214	150	56	
215-264	123	65	
≥ 265	136	54	
Multivitamins ^d			0.36
No	391	196	
Yes	211	91	

Abbreviations: kg=kilogram, m=meter, µg=microgram

^aP-values were calculated from a Fisher's exact test for categorical variables and a Student's t-test for numerical variables. ^bTotal number of drinks during the first trimester. ^cIncludes index child. ^dDuring month before pregnancy and first two months of pregnancy.

*Nominally associated (p<0.05) with maternal smoking status during pregnancy (smokers versus non-smokers).

Table 2. Characteristics of infants in the study according to maternal smoking.

Characteristic	Non-smokers	Smokers	P-value^a
All Infants	602	287	
Sex			0.25
Female	252	132	
Male	350	155	
Facial cleft status			1.8x10 ^{-4*}
Control	347	126	
Cleft lip with or without cleft palate	160	113	
Cleft palate only	95	48	
Gestational age (weeks)	38.4 ± 7.7	37.2 ± 9.9	0.06
Birth weight (kg)	3.6 ± 0.6	3.4 ± 0.7	3.4x10 ^{-6*}

Abbreviations: kg=kilogram

^aP-values were calculated from a Fisher's exact test for categorical variables and a Student's t-test for numerical variables.

*Nominally associated (p<0.05) with maternal smoking status during pregnancy (smokers versus non-smokers).

Table 3. Novel CpG sites confirmed at the site or gene-level.^a

Probe	Chr	Basepair (hg19)	Nearest gene	Model 1 Coeff ^b	Model 1 P-value	Replication: 450K ^c P-value
cg25189904	1	68299493	<i>GNG12</i>	-0.02	9.06E-09 [*]	4.96E-06
cg26764244 ^d	1	68299511	<i>GNG12</i>	-0.023	6.68E-10 [*]	4.47E-06
cg18703066 ^d	2	105363536	<i>LOC284998</i>	-0.006	7.63E-12 [*]	3.42E-07
cg18096987	3	11623873	<i>VGLL4</i>	-0.018	4.10E-08 [*]	2.20E-06
cg20344448 ^d	10	14372431	<i>FRMD4A</i>	0.015	2.97E-09 [*]	
cg11813497 ^d	10	14372879	<i>FRMD4A</i>	0.024	6.85E-09 [*]	3.48E-06
cg25464840 ^d	10	14372910	<i>FRMD4A</i>	0.022	3.15E-15 [*]	
cg15507334 ^d	10	14372913	<i>FRMD4A</i>	0.02	2.43E-13 [*]	
cg00029284 ^d	12	111731203	<i>CUX2</i>	-0.009	9.62E-06	1.53E-06
cg08698721 ^d	14	101294147	<i>MEG3</i>	0.013	1.66E-05	2.92E-06
cg04291079 ^d	14	101294430	<i>MEG3</i>	0.013	1.67E-07	
cg00253658 ^d	16	54210496	<i>FTO</i>	0.059	1.37E-17 [*]	9.64E-07
cg26681628	16	54210550	<i>FTO</i>	0.023	4.05E-13 [*]	
cg03687532	16	54228358	<i>FTO</i>	0.012	2.41E-05	
cg07339236 ^d	20	50312490	<i>ATP9A</i>	-0.013	2.21E-18 [*]	1.38E-07
cg16517298	1	230413174	<i>GALNT2</i>	-0.027	6.09E-06	
cg19727396 ^d	1	230415185	<i>GALNT2</i>	-0.018	1.15E-05	
cg24591105 ^d	1	230415225	<i>GALNT2</i>	-0.024	1.26E-07 [*]	
cg00589617 ^d	1	230415343	<i>GALNT2</i>	-0.029	1.07E-07 [*]	
cg05697274 ^d	1	230415377	<i>GALNT2</i>	-0.026	1.75E-08 [*]	
cg24250902 ^d	1	230415547	<i>GALNT2</i>	-0.032	1.13E-08 [*]	
cg03144619 ^d	1	230415668	<i>GALNT2</i>	-0.037	2.17E-09 [*]	
cg09368188 ^d	1	245330018	<i>KIF26B</i>	0.026	1.54E-07	

Abbreviations: Chr: chromosome, Coeff: coefficient

^aSite/gene region was not part of the 26 sites/10 gene regions that met genome-wide significance in Joubert et al. 2012 (Bonferroni correction, $p < 1.06 \times 10^{-7}$) but was confirmed using their results containing 74 additional nominally significant sites/69 gene regions. ^bA negative coefficient indicates reduced methylation in infants whose mothers smoked. ^cReplication at the site level using a previously published study (Joubert et al. 2012) that utilized Illumina HumanMethylation450 BeadChips. Note that those sites without direct confirmation were included if they were genome-wide significant in our study and a different CpG site in the same

gene region was identified in Joubert et al. 2012 (e.g. *GALNT2*). ^dMeets our criteria for dose-response (see Methods section, Tobacco smoke exposure, for details).

*Meets bonferroni correction ($p < 1.4 \times 10^{-7}$).

Figure legends

Figure 1. Manhattan plot of epigenome-wide results based on Model 1. The first grey horizontal line marks genome-wide significance using FDR $q < 0.05$ as a cutoff ($N=185$ CpGs), while the second grey horizontal line marks genome-wide significance using a Bonferroni correction ($N=68$ CpGs, $p < 1.40 \times 10^{-7}$). Blue crosses represent CpG sites that met genome-wide significance (FDR $q < 0.05$) and showed decreased methylation in infants whose mothers smoked. Red triangles represent CpG sites that met genome-wide significance and showed increased methylation in infants whose mothers smoked.

Figure 2. Replication of previously identified maternal smoking-related CpG sites (A) and confirmation of novel findings (B). (A) Twenty-six GWS CpG sites (Bonferroni correction; $p < 1.06 \times 10^{-7}$) related to maternal smoking during pregnancy were previously identified. Using our 185 GWS CpG sites (FDR $q < 0.05$) related to maternal smoking during pregnancy we further confirmed all 5 CpGs that had been previously replicated using a separate study population (Bonferroni correction for 26 tests; $p < 0.0019$) as described in Joubert et al. 2012. Further, we replicated 17 of the 21 CpGs that were not directly replicated in their small replication population using a Bonferroni correction for 26 tests. In total, 22 CpGs (6 unique gene regions) out of the 26 CpGs (10 unique gene regions) were either further confirmed or replicated at the site-level. An additional 17 CpGs, located in 7 of the 10 previously implicated or established gene regions, met genome-wide significance in our study, bringing the total to 39 CpGs in 7 gene regions; 22 CpGs replicated at the site-level and 17 CpGs replicated at the gene-level. (B) Novel findings included those CpG sites in gene regions not previously identified as shown in panel A. Publically-available results (Joubert et al., 2012) were used to confirm our findings. We considered two scenarios – the exact CpG was confirmed versus the exact CpG was not

confirmed, but other CpGs were identified in the same gene region. Abbreviations: GWS: genome-wide significant; N: number

Figure 1

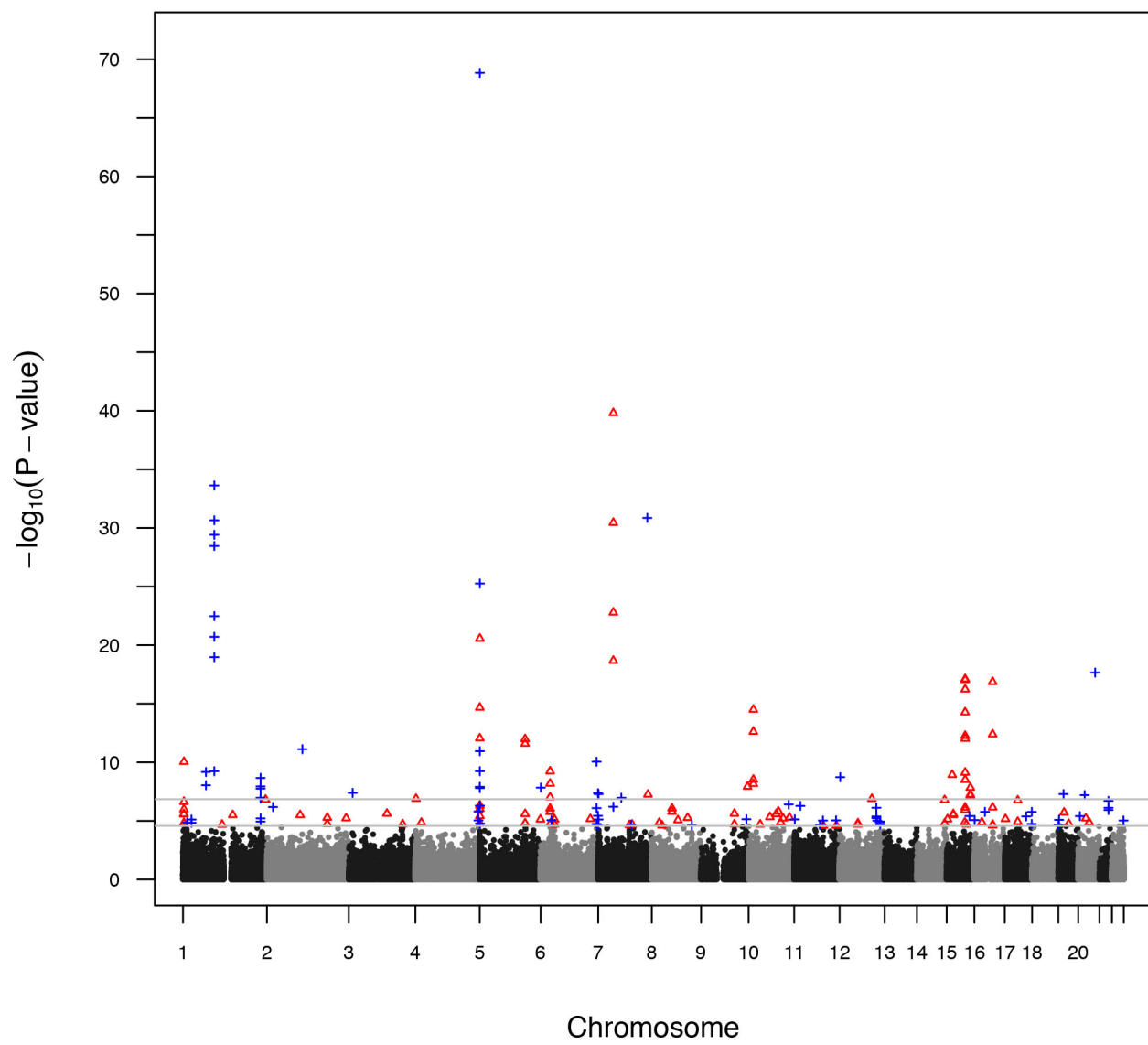


Figure 2

A) Previously identified gene regions related to maternal smoking during pregnancy

<u>Categories</u>	<u>Joubert et al. 2012</u>		<u>Our GWS findings</u>
Established sites	5 CpGs	Further confirmed →	5/5 CpGs
Implicated sites	21 CpGs	Replicated →	17/21 CpGs
	<div style="border-top: 1px solid black; width: 100px; margin: 5px auto;"></div> 26 CpGs previously identified Total of 10 unique gene regions		<div style="border-top: 1px solid black; width: 100px; margin: 5px auto;"></div> 22 CpGs replicated Total of 6 unique gene regions

B) Novel findings

b) Novel findings

146 novel sites
(103 gene regions)

Site confirmed → 9 CpGs } 8 gene regions

Different site, same gene region → 2 gene regions

Gene region	N GWS sites	N Confirmed sites
GNG12	2	2
LOC284998	1	1
VGLL4	1	1
FRMD4A	4	1
CUX2	1	1
MEG3	2	1
FTO	3	1
ATP9A	1	1

Gene region	N GWS sites	N Confirmed sites
GALNT2	7	0
KIF26B	1	0